The selection of high-impact health informatics literature: a comparison of results between the content expert and the expert searcher

Elizabeth C. Whipple, MLS; Julie J. McGowan, PhD, FACMI, AHIP, FMLA; Brian E. Dixon, MPA; Atif Zafar, MD

See end of article for authors' affiliations.

Background: The Agency for Healthcare Research and Quality (AHRQ) National Resource Center for Health Information Technology (NRC) created the Health IT Bibliography that contains peer-reviewed articles in eleven different health informatics categories. To create the bibliography, informatics experts identified what they considered the seminal articles in each category.

Methods: Using the same eleven categories, an expert searcher (librarian) compiled a list of the ''best'' health informatics articles using information seeking and retrieval tools. The two sets of articles were then compared using high citation counts as a measure of value.

Results: The expert searcher set (8,230) contained more than 3 times the citations to chosen articles compared to the content expert set (2,382). Of 60 articles, 27% of those articles (n=16) were included in both sets. The frequently cited journals were similar for both sets, and one-third of the same authors were cited in both sets.

Discussion: While citation counts and the timeliness of the articles differed in the two sets, the same authors and same journals were frequently present in both sets.

Conclusion: A best practice for locating high-quality articles may be collaboration between expert searchers and content experts.

INTRODUCTION

Determining the best articles on a particular subject is arguably a subjective process that depends in large part on the intention of the query. Many parameters can be used to determine whether or not an article is integral to a particular field. One of the most common parameters is an article's citation rate: the number of times a particular article is cited by other papers, presentations, and conference proceedings. The more often a paper is cited, the greater its impact in the field of scientific discovery and innovation. The purpose of this study was to compare the best peer-reviewed articles chosen by health informatics experts to the best peer-reviewed articles chosen by a librarian, using citation rate to measure an article's value.

While librarians, as expert searchers, are a natural fit for locating and selecting the best articles, experts in a particular subject or discipline are also obvious candidates for such a task. As invested researchers in a particular field of study, subject experts would be familiar with the seminal papers and important researchers in their own particular fields. The study reported here stems from a project done for the Agency for Healthcare Research and Quality (AHRQ), which produced a bibliography of the best peer-reviewed articles on several subjects in the field of health information technology (IT) as determined by content experts.

Highlights

- The expert searcher found articles that received a greater number of citations than the content experts overall and significantly more in several categories. However, overall, content experts selected more current articles.
- Two independently derived sets of high-impact health informatics articles overlapped only 27% (n=16).

Implications

- Although the process of finding "the best" articles for a given discipline is somewhat subjective, there are several accepted methods for selecting top articles. A best practice for creating bibliographies of top articles is using the combined knowledge and skills of expert searchers, who, as information science professionals, can identify relevant and high-quality articles using proven techniques and tools, and content experts, who, as domain professionals, can refine article sets using subject expertise and acumen.
- These two different methodologies produced very different sets of high-impact articles; collaboration between content experts and expert searchers is ideal.

In 2006, the AHRQ National Resource Center for Health Information Technology (NRC) created an online knowledge repository of AHRQ and non-AHRQ resources that emphasize best practices for the adoption and use of health informatics applications, such as electronic health record (EHR) systems.

This article has been approved for the Medical Library Association's Independent Reading Program <http://www.mlanet.org/education/irp/>.

This repository contains a variety of items, including a sample request-for-proposal document for use in selecting a vendor, a market assessment of open source ambulatory EHR systems, and a toolkit for evaluating health information technologies. The repository and website provide health care professionals and organizations with knowledge resources to support greater adoption and use of health IT applications across the United States. Since its creation, AHRQ's online knowledge repository has grown and now includes more than 7,000 items. To assist users in finding the information they seek, the NRC has invested in the development of search tools and interactive user interfaces. Basic search tools and interfaces, however, may not be the most appropriate tools for novice users to successfully find the information they seek [1].

To better guide users to targeted items in its knowledge repository, the NRC created the Health IT Bibliography [2]. The NRC invited informatics experts to identify implementation-focused resources from the peer-reviewed literature and NRC knowledge repository. Experts were asked to suggest resources they felt would be helpful to other health care professionals seeking to develop, purchase, implement, and use health IT in the routine care of patients in hospitals, physician offices, and other settings, such as nursing homes (Figure 1). The NRC then organized the selected resources into eleven broad health informatics categories (Table 1).

The selected categories were prioritized based on their alignment with AHRQ objectives and areas of interest. When the bibliography was developed, AHRQ supported a broad portfolio of health IT projects focused primarily on EHRs, clinical decision support systems (CDSS), computerized provider order entry (CPOE), electronic prescribing (eRx), and health information exchange (HIE) [3]. AHRQ also has a long history of supporting patient safety research [4, 5] and efforts to create standards for interoperability between health IT systems [6]. Although experts were asked to identify peer-reviewed and non-peer-reviewed resources for the bibliography, this paper discusses only the selection of peer-reviewed articles.

Librarians are not strangers to searching the literature and finding the best articles. Librarians are information professionals, skilled in understanding and utilizing information management seeking tools to conduct expert searches [7–11]. McKibbon et al. noted that librarians conducting MEDLINE searches had significantly better recall and precision rates than content experts who were novice searchers and had recall equivalent to—and better precision than—experienced MEDLINE users [7]. Davidoff and Florance have promoted the ''librarian as informationist'' model, because physicians have not been trained in the same information retrieval skills as librarians [9]. While the literature indicates that librarians can be expert searchers, physicians do not always agree. Arnott Smith noted that only 25% of health care professionals (mostly physicians) ''be-

lieved that a librarian could find all relevant research articles required to support their evidence-based practice'' [10].

In the case of this collection of health informatics topics, we were interested to know which articles a librarian, as an expert searcher, would determine to be best. The purpose of this study was to compare the best peer-reviewed articles (in the eleven categories) chosen by health informatics experts (henceforth called content experts) to the best peer-reviewed articles chosen by a librarian (henceforth called expert searcher). The comparison was to be made on the basis of the number of times articles from each set have been cited, indicating the impact and influence of these articles in the health informatics field.

METHODS

Expert searchers have many information retrieval and evaluation tools at their disposal. This study used PubMed searching, the Journal Citation Reports (JCR), publication type limits (e.g., review articles), and citation analysis tools (including the ''Cited Reference'' feature from ISI).

The Health IT Bibliography covers eleven different health informatics topics. For each of these predetermined topics, PubMed MEDLINE was searched using a combination of Medical Subject Heading (MeSH) terms and relevant keywords for each topic. After these initial searches, the ISI's 2006 JCR Science Edition was employed to identify the top journals in the discipline of medical informatics. Twenty journals were included in the JCR Medical Informatics category, ranked by impact factor. The PubMed searches were first limited to the top ten journals (based on impact factor). If enough relevant results were retrieved, articles from only those top ten journals were used. If the retrieval set was less than five articles, the search was expanded to include articles from the top fifteen or top twenty journals. Limiting to these ten (or fifteen, or twenty) journals was intended to retrieve articles that would be cited by articles published in the most prestigious medical informatics journals.

Aside from using MeSH terms, keywords, and top-ranked journals to create a retrieval set from which to determine the best articles, limiting to review articles was used to further refine the retrieval set. Review articles provide an overview of a topic, the main researchers involved in that topic, and the current state of research. Ideally, review articles contain citations to the most influential papers in a particular area of research. Searches were limited to review articles as often as possible; however, as in the case of limiting to the top ten (or fifteen, or twenty) JCR medical informatics journals, if the retrieved results were not relevant or contained too few citations, the search was expanded to non-review articles. Using these multiple methods—MeSH terms and keywords, limits to top-ranked journals, and limits to review articles—sets of articles were generated to locate the

best articles for each health informatics category (Figure 2).

Using ISI's Web of Knowledge, the items cited by each set of articles were retrieved in machine readable form and placed in an Excel spreadsheet. The spreadsheet was sorted to determine which had been cited the most frequently and therefore would likely have the greatest impact and represent the best articles on that health informatics topic. This procedure was followed for each of the eleven categories, and the same number of best articles selected as for the corresponding Health IT Bibliography (e.g., the bibliography had eight articles on CDSS, so the eight CDSS articles that were most frequently cited in the papers retrieved in the expert searcher's CDSS search were selected) (Table 1).

To compare these expert searcher sets with the content expert sets, the "Cited Reference" feature of ISI's Web of Knowledge was used to determine the number of times each article in these sets and in the Health IT Bibliography had been cited. To better understand differences between the two sets, data were also collected on the journals in which these articles were published, the number of authors represented in each of the sets, and the overlap in authors and articles between the expert searcher and the content expert sets.

RESULTS

The Health IT Bibliography had 60 articles altogether, and therefore 60 articles were chosen for the expert searcher article sets. From both sets, 16 articles (27%) were the same. In the adoption strategies, standards and interoperability, and workflow analysis catego-

ries, the content expert articles were cited more often than were the expert searcher articles. In the other eight categories, the expert searcher articles were cited more than the content expert articles (Figure 3).

The highest number of citations for the content expert set was in the following categories: patient safety (594), CDSS (425), adoption strategies (398), CPOE systems (251), and standards and interoperability (212). The highest number of citations for the expert searcher set occurred for a different group of categories: business case (2,283), patient safety (1,618), CDSS (1,032), eRx (878), and CPOE systems (776).

The expert searcher's 6 most highly cited sets were each cited more than the highest content expert set. Overall, for the total number of citations for each set, the expert searcher set (8,230) contained more than 3 times the citations to articles chosen from the content expert set (2,382).

Table 1

Health Informatics categories from the Agency for Healthcare Research and Quality (AHRQ) National Resource Center for Health Information Technology (NRC) Health IT Bibliography along with the number of peer-reviewed articles in each category

Health informatics category	Articles
Adoption strategies	6
Business case	6
Clinical decision support systems (CDSS)	8
Computerized provider order entry (CPOE) systems	5
Electronic health record (EHR) systems	5
Electronic prescribing (eRx)	4
Health information exchange (HIE)	7
Standards and interoperability	4
Evaluation studies in health information technology (IT)	7
Patient safety	6
Workflow analysis	2

Figure 1

Process for content experts choosing articles for the Agency for Healthcare Research and Quality (AHRQ) National Resource Center for Health Information Technology (NRC) Health Information Technology (IT) Bibliography
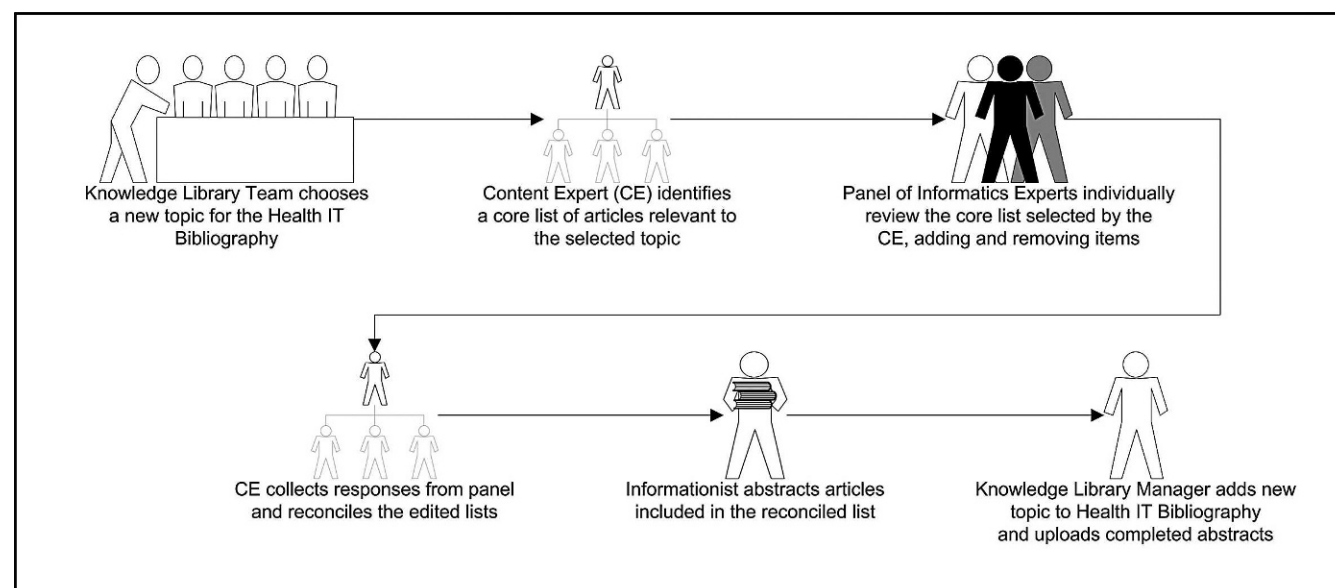
Figure 2
Process for expert searcher choosing health informatics articles
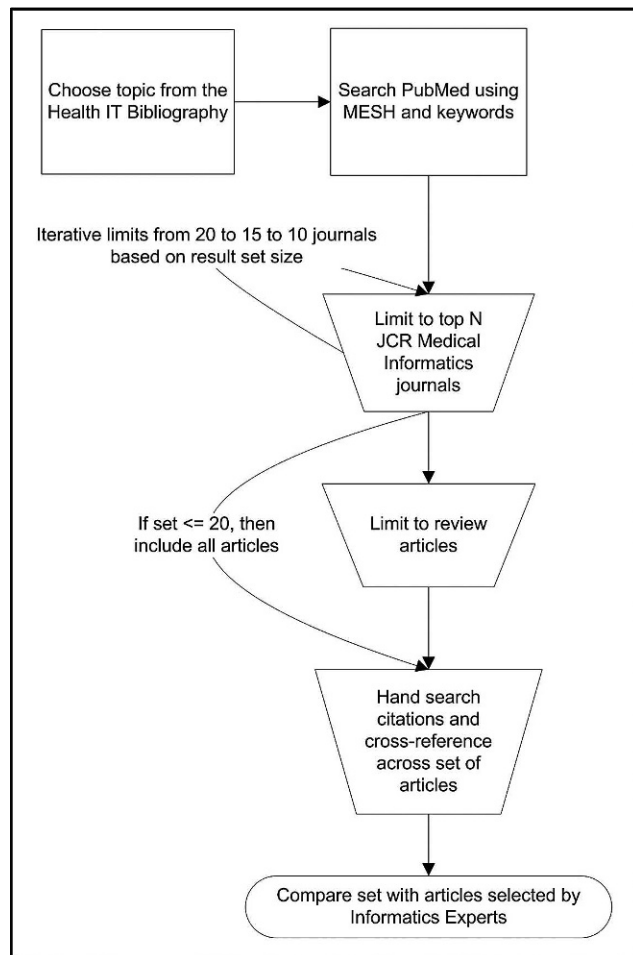
The categories in which the number of citations were the most similar for both sets include: workflow analysis (15 citation difference), EHR (96 citation difference), adoption strategies (98), standards and interoperability (117), and evaluation studies in health IT (121).

Publication date for articles in both sets ranged from 1991–2007. Content experts chose the majority of articles from those published in the past 5 years (88% were published from 2002–2007), while the expert searcher set had a more even distribution of articles across the time frame (only 50% from 2002–2007) (Figure 4).

The frequently cited journals were similar between the 2 groups. When journals were ranked in terms of frequency of citation, the top 6 for the content experts were *Journal of the American Medical Informatics Association (JAMIA)* (21), *Health Affairs* (8), *Annals of Internal Medicine* (4), *International Journal of Medical Informatics* (3), *Journal of the American Medical Association (JAMA)* (3), and *Journal of Biomedical Informatics* (3). Four out of the 6 of these top journals were included in the expert searcher group: *JAMIA* (19), *JAMA* (10), *Health Affairs* (6), *International Journal of*

Medical Informatics (4), *Archives of Internal Medicine* (3), and *New England Journal of Medicine* (3). *JAMIA* was cited the most in both sets of articles; *JAMIA* also has the highest impact factor for the group of Medical Informatics journals (3.979). The *International Journal of Medical Informatics*, ranked fourth and third, respectively, is also in the JCR Medical Informatics category. *Health Affairs* also ranks highly in both sets (second and third, respectively), although this journal is not included in JCR's Medical Informatics category (it is ranked fifth in the Health Care Sciences and Services JCR category).

The order in which authors are listed on a paper generally denotes the extent to which they contributed to the paper, although authorship may rotate between members of a research team and sometimes the last name listed is the most important (often the head of the lab group). Looking at all authors for all papers, the content expert set had 250 unique authors, while the expert searcher set had 258 unique authors. Thirty-eight authors were cited more than once in the content expert set, compared with thirty-seven authors in the expert searcher set. Twelve authors were cited more than once in both sets. Authors cited in both sets (n=101) made up 40.4% of authors in the content expert and 39.1% in the expert searcher set, respectively (Table 2).

Timeliness of articles can be an important consideration, especially with respect to those that deal with technology. Because technology continues to change at a rapid rate, technological issues in the literature 10 years ago may not be relevant to discussion of technology in more recent articles. Focusing on recently published articles can be important to researchers: Articles from the period 2002–2007 account for over three-fourths (88%) of the content expert set, while articles from that period account for only half (50%) of the expert searcher set. One would expect that the number of citations to articles in the bigger set (the content expert set) would be greater than in the smaller expert searcher set. The number of citations to the articles in those 2002–2007 sets, however, is 1,604 (content expert) and 2,008 (expert searcher), respectively. The expert searcher set included fewer articles published in the past 6 years than the content expert set, but for those articles chosen, the amount of citations for the expert searcher set was 25% more than for the content expert set.

The journal's impact factors for articles chosen in each set ranged from 1.068–51.296. Impact factors present an indication of how often a journal is cited. For this comparison, however, journal impact factor did not by itself predict the articles selected for either set.

The authors hypothesized that the articles in the content expert set would reflect the important authors in the field and that an author analysis would show a majority of the articles grouped around a few authors and some outliers, while the expert searcher set would be more uniformly spread out over many authors. In fact, the author comparison results between the 2 sets did not differ much at all. The

Figure 3
Number of times articles in each health IT category were cited
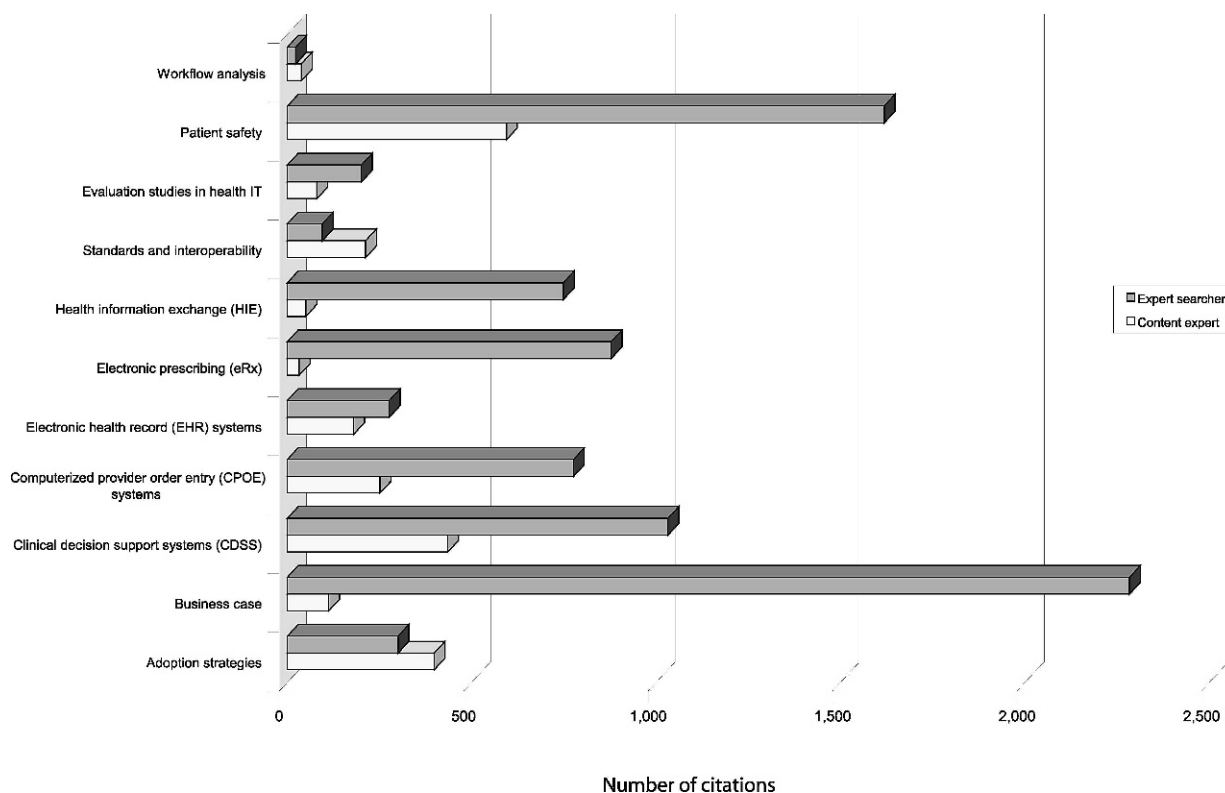
Figure 4
Chosen articles published by year
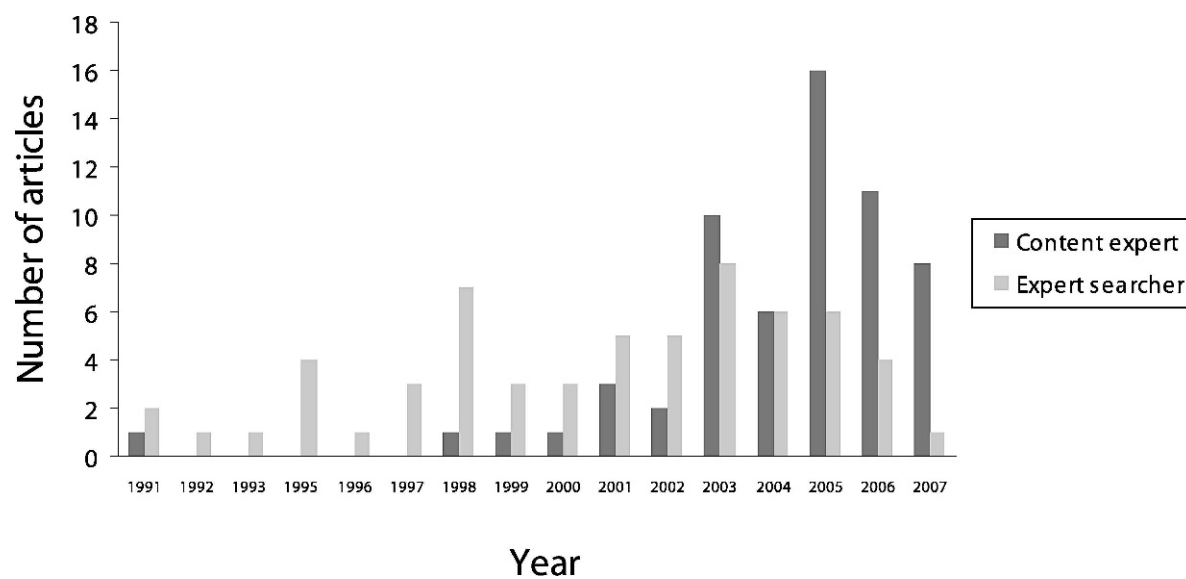
Table 2
Comparisons of authors from the content expert and expert searcher sets

Author comparisons	Content expert set	Expert searcher set
Number of individual authors	250	258
Number of authors cited more than once	38	37
Authors cited at least once in both sets	101	101
Percentage of authors cited in both sets	40%	39%

category ''authors cited more than once in a set,'' which would produce grouping, only occurred 15.2% and 14.3% of the time in the content expert set and expert searcher set, respectively. Additionally, 71.4% of the authors cited more than once in the content expert set were also cited in the expert searcher set, and 64.9% of the articles cited more than once in the expert searcher set were also cited in the content expert set. A difference in the authors that had been expected to emerge between the 2 sets never did: The sets cited the same authors more than one-third of the time and had the same amount of clustering around authors.

DISCUSSION

This paper assumes that the number of times an article is cited indicates its impact on the scientific field; however, this is certainly not the only measure of the quality or importance of an article.

Sixteen articles (27%) appeared in both the content expert and expert searcher sets. Of those 16, 7 were used for the same subject; 9 of the 16 were used in both sets, but in different health informatics categories. The reason for this discrepancy relates to the category topics not being mutually exclusive (for instance, an article on EHR systems in the content expert set was used as a business case article in the expert searcher set). An article that mentions HIE and interoperability could belong to the HIE, standards and interoperability, or business case categories. The decision as to where an article best fits may be a matter of personal opinion.

Evidence that many of these categories overlap and interrelate is also apparent in author citations. An author may publish articles on HIE, evaluation studies in health IT, CDSS, patient safety, and eRx (this example from the content expert set). While differences exist among the health informatics categories, the amount of overlap between interrelated topics (and therefore an absence of clear-cut lines between some of the health informatics categories), and hence content experts' article choices, is not surprising.

The JCR can also be a helpful starting point for identifying medical informatics journals. However, out of the twenty journals listed in the Medical Informatics category, only five of them contained articles in either set. *Health Affairs*, a journal cited more than once in both sets, is in the JCR Health Care Sciences and Services category, not the Medical Informatics category. Even if the main focus of *Health*

Affairs is not health informatics, both sets of article agree that important health informatics articles are being published in this journal.

While results between the content expert set and the expert searcher set have some similarities (a quarter of the same articles, similar journals, and more than a third of the same authors), great variation existed between the two sets. In spite of those variations, both sets could be argued to be representative of the best articles in health informatics. Overall, the expert searcher set had more citations than the articles the content experts chose. Whether or not the articles chosen by the expert searcher are, in fact, better articles may be a source of further investigation.

Limitations of this study

Given the cross-disciplinary nature of the field of informatics, searching PubMed exclusively might have limited the possible results for this study. A further exploration of other databases that contain informatics articles would contribute to the final findings of this study. Limiting to the JCR Medical Informatics category might also have limited the retrieval possibilities, and there might be better ways of limiting retrieval sets. However, in spite of this choice, as noted above, *Health Affairs*, a journal not in the JCR Medical Informatics category, was highly cited in both the content expert and expert searcher sets. In addition, this study was limited to only peer-reviewed articles, while recognizing that the AHRQ Health IT Bibliography contains both peer-reviewed and non-peer-reviewed materials and that informaticists might rely on both types of articles when conducting research.

CONCLUSION

Many different measurements and criteria can determine whether an article is ''the best'' in a particular field. Some of the more easily quantifiable measurements are how often the article is cited, in which journal it appears, who the authors are, and how recently it was published. According to the comparisons in this article, articles in the expert searcher's set of articles had more impact than articles selected by the content experts—if impact is judged by looking at the one criterion of the number of times the articles have been cited. Expert searchers, while not necessarily having content expertise in a particular topic, have tools at their disposal that can prove to be a valuable asset to determine the best articles in an area. Having the training and background in understanding database organization and the underlying information architecture for many of the information systems currently available make expert searchers ideal candidates to find relevant information, despite lacking subject expertise. Subject expertise certainly provides a broader background from which to draw, but subject expertise alone is not the only method to determine relevant and useful articles. Conversely, while expert searchers are certainly not replacing

content experts, collaborating with informationists (i.e., expert searchers) for research information needs can prove to be a synergistic relationship.

REFERENCES

1. Hölscher M, Strube G. Web search behavior of Internet experts and newbies. Intl J Comput Telecomm Netw. 2000 Jun;33(1–6):337–46.
2. Agency for Healthcare Research and Quality, National Resource Center for Health Information Technology. Health IT bibliography: top resources for key topics [Internet] The Center [rev. 13 Mar 2008; cited 21 Aug 2008]. <http://www.healthit.ahrq.gov/portal/server.pt?open=512&objID=653&&PageID=12790&mode=2&in_hi_userid=3882&cached=true>.
3. Dixon BE. The landscape of the AHRQ Health Information Technology Portfolio. AMIA Annu Symp Proc. 2006:912.
4. Romano PS, Geppert JJ, Davies S, Miller MR, Elixhauser A, McDonald KM. A national profile of patient safety in U.S. hospitals. Health Aff (Millwood). 2003 Mar–Apr;22(2):154–66.
5. Agency for Healthcare Research and Quality. AHRQ quality indicators—patient safety indicators: software documentation. Rockville, MD: The Agency; 2002.
6. Fitzmaurice JM, Adams K, Eisenberg JM. Three decades of research on computer applications in health care: medical informatics support at the Agency for Healthcare Research and Quality. J Am Med Inform Assoc. 2002 Mar–Apr; 9(2):144–60.
7. McKibbon KA, Haynes RB, Dilks CJ, Ramsden MF, Ryan NC, Baker L, Flemming T, Fitzgerald D. How good are clinical MEDLINE searches? a comparative study of clinical end-user and librarian searches. Comput Biomed Res. 1990 Dec;23(6):583–93.
8. Task Force on Expert Searching, Medical Library Association. Medical Library Association policy statement: role of expert searching in health sciences libraries [Internet]. Chicago, IL: The Association [Sep 2003; cited 7 Sep 2008]. <http://www.mlanet.org/resources/expert_search/policy_expert_search.html>.
9. Davidoff F, Florance V. The informationist: a new health profession? Ann Int Med. 2000 Jun;132(12):996–8.
10. Arnott Smith C. An evolution of experts: MEDLINE in the library school. J Med Libr Assoc. 2005 Jan;93(1):53–60.
11. McLellan F. 1966 and all that—when is a literature search done? Lancet. 2001 Aug 25;358(9282):646. DOI: 10.1016/S0140-6736(01)05826-3.

AUTHORS' AFFILIATIONS

Elizabeth C. Whipple, MLS, ewhipple@iupui.edu, Research Informationist/Assistant Librarian, Ruth Lilly Medical Library; **Julie J. McGowan, PhD, FACMI, AHIP, FMLA,** jjmcgowa@iupui.edu, Associate Dean, Information Resources and Educational Technology, and Professor, Knowledge Informatics and Pediatrics; Indiana University School of Medicine, 975 West Walnut Street, IB-310, Indianapolis, IN 46202-512; **Brian E. Dixon, MPA,** bdixon@regenstrief.org, Health Information Project Manager, Regenstrief Institute, 410 West 10th Street, Suite 2000, Indianapolis, IN 46202-3012; **Atif Zafar, MD,** azafar@iupui.edu, Associate Professor, Department of Medicine, Indiana University School of Medicine, 410 West 10th Street, Suite 2000, Indianapolis, IN 46202-3012